

## Computer Analyses of Characteristic Infrared Bands of Globular Proteins

MAX RÜEGG,\* VIRGINIA METZGER, and H. SUSI,† *Eastern Regional Research Center, ARS, U.S. Department of Agriculture, Philadelphia, Pennsylvania 19118*

### Synopsis

Infrared spectra of myoglobin, ribonuclease, lysozyme,  $\alpha$ -chymotrypsin,  $\alpha$ -lactalbumin, and  $\beta$ -lactoglobulin A were obtained in deuterium oxide solution in units of absorbance versus wavenumber from 1340 to 1750  $\text{cm}^{-1}$ . The spectra were resolved into Gaussian components by means of an iterative computer program. Resolved characteristic absorption peaks for the two infrared active amide I' components of antiparallel chain-pleated sheets ( $\beta$ -structure) were obtained. The characteristic amide I' peaks of  $\alpha$ -helical regions and apparently unordered regions overlap in  $\text{D}_2\text{O}$  solution. Absorptivity values for the resolved  $\beta$ -structure peak around 1630  $\text{cm}^{-1}$  were estimated on the basis of the known structure of ribonuclease, lysozyme, and  $\beta$ -chymotrypsin. The  $\beta$ -structure content of  $\beta$ -lactoglobulin was estimated to be ca. 48%, of  $\alpha$ -lactalbumin ca. 18%, and of  $\alpha_s$ -casein close to zero. The results are in general agreement with conclusions drawn from circular dichroism and optical rotatory dispersion studies.

### INTRODUCTION

Secondary amides, polypeptides, and proteins exhibit characteristic infrared absorption bands around 1650 and 1550  $\text{cm}^{-1}$ , which are usually called the amide I and II bands, respectively.<sup>1-3</sup> If the peptide NH group is deuterated, the second band, associated largely with NH or ND bending vibrations,<sup>1,2</sup> is shifted to ca. 1450  $\text{cm}^{-1}$ . The two bands are now called amide I' and II'.<sup>2</sup> The exact positions of the band centers, and the observable fine structure, if any, can be used to draw conclusions about the secondary structure of the polypeptide or protein under investigation.<sup>2,4-6</sup> Extensive studies have been carried out on solid fibrous proteins.<sup>4</sup> A number of globular proteins have been studied in  $\text{D}_2\text{O}$  solution<sup>5,6</sup> where the NH grouping is converted to ND. Unfortunately, only the amide I' band can be analyzed with any reliability in  $\text{D}_2\text{O}$  solution because traces of HOD absorb very close to the amide II' frequency and distort the 1450- $\text{cm}^{-1}$  region of the spectra.<sup>5</sup>

Although it is sometimes possible to conclude from infrared spectra that a given protein has, for instance, a large amount of  $\beta$ -structure,<sup>7</sup> or contains segments with different secondary structures,<sup>4</sup> such informa-

\* Present address: Federal Dairy Research Station, 3097 Liebfeld-Bern, Switzerland.

† To whom correspondence should be addressed.

tion is highly qualitative.<sup>4-7</sup> In order to find out whether information of a more quantitative nature can be gained from infrared studies of proteins, we have analyzed the 1300–1750-cm<sup>-1</sup> region of D<sub>2</sub>O solution spectra of several globular proteins in some detail, by employing a computer program that resolves the observed spectra into individual absorption bands. Although solid-state infrared spectra sometimes yield more detailed information than solution spectra (because polarization measurements are possible in the solid state),<sup>2,4</sup> they are not well suited for semiquantitative studies because the thickness of solid protein films is neither readily determined nor reproduced. In addition, secondary effects caused by refraction can distort the observed band envelopes. In D<sub>2</sub>O solution, the shape and fine structure of the amide I band (~1630–1680 cm<sup>-1</sup>) can be studied in some detail while the amide II' band (~1450 cm<sup>-1</sup>) is distorted by HOD absorption.

## EXPERIMENTAL

The spectra of the following proteins were investigated in D<sub>2</sub>O solution: myoglobin, ribonuclease, lysozyme, chymotrypsin,  $\alpha_s$ -casein, native  $\beta$ -lactoglobulin A, and alkali-denatured  $\beta$ -lactoglobulin A. The  $\beta$ -lactoglobulin was obtained from Dr. R. E. Townend of this Center and had been twice recrystallized, sample "cow 5215."  $\alpha$ -lactalbumin was obtained through the courtesy of Prof. M. Kronman (New York Upstate Medical Center, Department of Biochemistry, Syracuse, N.Y.).  $\alpha_s$ -casein was obtained from Dr. H. M. Farrell of this Center. The other proteins were commercial preparations from the following sources: myoglobin, Sigma 3C-3440;  $\alpha$ -chymotrypsin, Worthington CD12DC; ribonuclease, crystallized, Mann; lysozyme, Sigma lot 14C-8270. The crystal structure of myoglobin,<sup>8</sup> ribonuclease,<sup>9</sup> lysozyme,<sup>10</sup> and  $\alpha$ -chymotrypsin<sup>11</sup> is known from X-ray investigations; the structure of the remaining three proteins is not.

The proteins were dissolved in D<sub>2</sub>O to yield 2% solutions by weight in the presence of 0.01 M sodium chloride. pH adjustments, if necessary, were carried out by adding NaOD or DCl in D<sub>2</sub>O solution. Sealed cells with calcium fluoride windows and 0.1-mm path length were employed to obtain infrared spectra. The sample was placed in the working beam, a matched cell filled with D<sub>2</sub>O in the reference beam of a double-beam spectrometer. The spectra were recorded from 1300 to 1750 cm<sup>-1</sup> in order to obtain a useful baseline, although only the amide I' band around 1650 cm<sup>-1</sup> can be used for studies of the secondary structure in D<sub>2</sub>O solution. A Beckmann model IR-7 double-beam grating spectrometer was employed to record differential spectra (solution versus solvent) directly in absorbance units versus wavenumber.

The obtained spectra were resolved into Gaussian components with a computer program written by C. Roland Eddy and Virginia Metzger of this Center (available on request). The program performs iterative

curve resolution based on direct search solutions.<sup>12</sup> The input information consists of the following: 1) the analytical nature (shape) of the resolved peaks; a Gaussian form was assumed and found to yield satisfactory agreement between observed and calculated curves (see below); 2) the number and the peak frequency range of resolved bands (i.e., the frequency range in which each peak is allowed to vary); 3) baseline points outside the absorption region where the total absorbance is initially assumed to be zero. The iteration procedure seeks the best agreement between observed and calculated curves by varying the slope and intercept of the linear baseline, the peak positions (frequencies), the peak heights (intensities), and the half-widths of the resolved peaks. Data points were supplied to the computer for every 2  $\text{cm}^{-1}$  in the 1760–1520- $\text{cm}^{-1}$  range, and every 4  $\text{cm}^{-1}$  for the range 1520–1340  $\text{cm}^{-1}$ . The data points were assigned weights proportional to absorbance to give more weight to intense bands. The output includes a listing of final peak positions, peak heights, half-widths and areas, and a plotted diagram showing the experimental curve, the calculated curve, and the individual resolved peaks. Figure 1a provides an example where a total of nine peaks were necessary to fit the experimental curve. Computations were carried out with an IBM model 1130 computer equipped with a Calcomp curve plotter.

## RESULTS AND DISCUSSION

### Ribonuclease

Figure 1a shows the observed and calculated absorption spectrum of ribonuclease from 1340 to 1750  $\text{cm}^{-1}$ , as well as the resolved individual peaks. The structure of this protein in the crystalline state is known.<sup>9</sup> The secondary structure, based on X-ray data, can be described as follows:<sup>13</sup> 44%  $\alpha$ -helix; 25% antiparallel chain-pleated sheet (also called  $\beta$ -structure); 31% unordered structure. The input data specified nine peaks, assigned to the following absorption bands:

- ~1670 and ~1630  $\text{cm}^{-1}$ ,  $\nu(0,\pi)$  and  $\nu(\pi,0)$  components of the amide I' band of antiparallel chain-pleated sheets ( $\beta$ -structure);<sup>4</sup>
- ~1650  $\text{cm}^{-1}$ , the  $\nu(0)$  component of the amide I' band of  $\alpha$ -helical segments<sup>4</sup> and the amide I' of the unordered segments as observed in  $\text{D}_2\text{O}$  solution;<sup>5</sup> these two bands cannot be resolved in  $\text{D}_2\text{O}$  solution on the basis of available data because the frequency difference is only a few wavenumber units;<sup>5</sup>
- ~1615  $\text{cm}^{-1}$ , unassigned;
- ~1585  $\text{cm}^{-1}$ , asymmetric stretching of side-chain  $\text{COO}^-$  groups;<sup>3</sup>
- ~1515  $\text{cm}^{-1}$ , unassigned;
- ~1450  $\text{cm}^{-1}$ , amide II' + HOD bending;<sup>14</sup>
- ~1400 symmetric stretching of  $\text{COO}^-$  groups.<sup>3</sup>

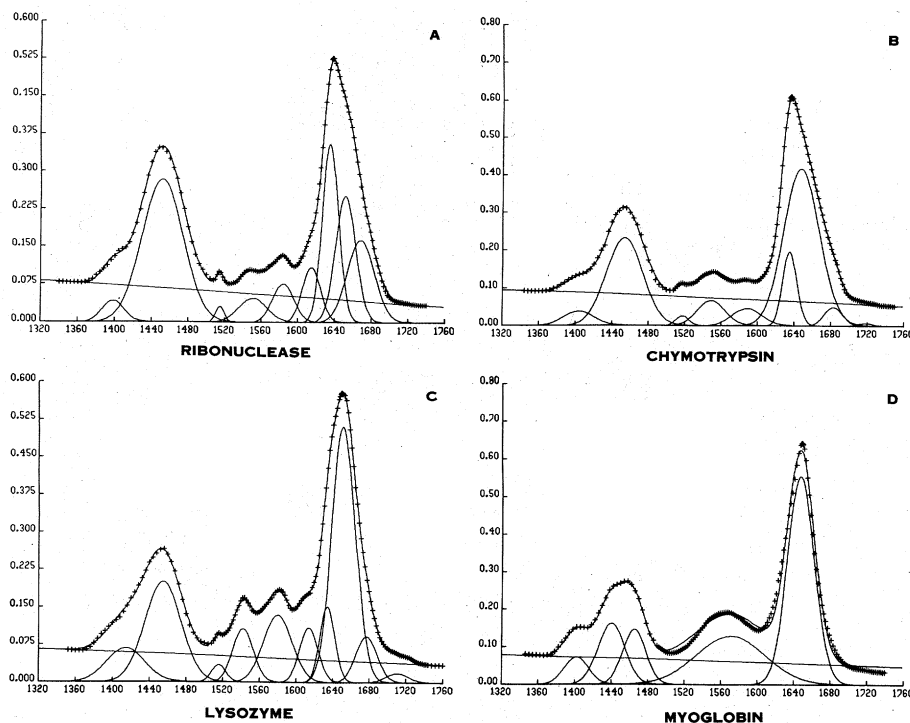


Fig. 1. Computer plots of observed (+) and calculated (—) infrared spectra. The pD values of the 2% protein solutions were: 5.0 (A), 4.2 (B), 5.2 (C), and 7.2 (D). Absorbance is plotted versus wavenumber; path length 0.1 mm.

The assignment of the weak bands between  $1500$  and  $1600\text{ cm}^{-1}$  is somewhat tentative, but does not interfere with the interpretation of the spectra in terms of different segments of the secondary structure. The peaks have to be included in the computations in order to obtain a realistic baseline and a correct presentation of the components of the amide I band around  $1650\text{ cm}^{-1}$ . It is evident on the basis of Figure 1a that the relative amounts of the  $\alpha$ -helical conformation and the unordered conformation cannot be determined from the intensity of the single peak around  $1650\text{ cm}^{-1}$ . On the other hand, the reasonably well-resolved  $1630\text{-cm}^{-1}$   $\beta$ -structure band, in conjunction with the known percentage of  $\beta$ -structure in this protein,<sup>9,13</sup> can be used to arrive at an absorptivity value for the resolved band, which can, in turn, be used to estimate the amount of  $\beta$ -structure in proteins of unknown conformation. Such estimations are attempted below. (The information obtainable from infrared studies is thus complementary to circular dichroism and optical rotatory dispersion investigations, which yield estimations regarding the  $\alpha$ -helix content of polypeptides and proteins.)

#### Myoglobin, Lysozyme, and Chymotrypsin

Figure 1b, c, and d show the observed and calculated infrared spectra of these three proteins with known conformation.<sup>8,10,11,13</sup> The band as-

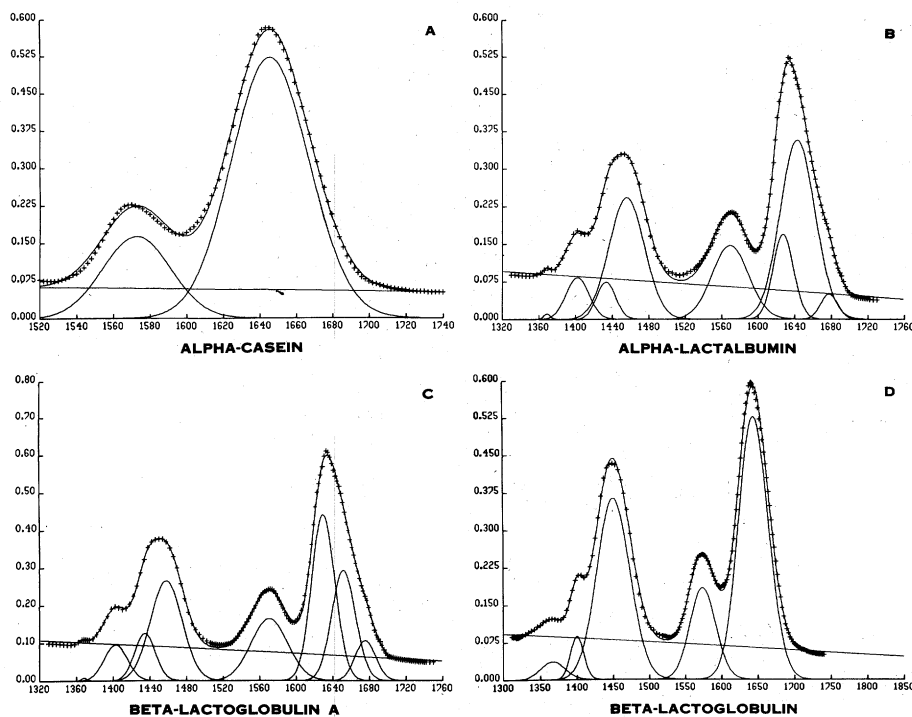


Fig. 2. Computer plots of observed (+) and calculated (—) infrared spectra of  $\alpha_s$ -casein (A),  $\alpha$ -lactalbumin (B),  $\beta$ -lactoglobulin-A (C), and alkali denatured  $\beta$ -lactoglobulin A (D). The pD values of 2% solutions were: 7.3 (A), 6.9 (B), 6.9 (C), and 11.9 (D). Absorbance is plotted versus wavenumber; path length 0.1 mm.

signments are quite similar to the discussed case of ribonuclease. In particular, the bands close to  $1680$  and  $1630\text{ cm}^{-1}$  can be assigned to the  $\nu(0,\pi)$  and  $\nu(\pi,0)$  components of the amide I' associated with the  $\beta$ -structure,<sup>4</sup> and the band close to  $1650\text{ cm}^{-1}$  to the  $\nu(0)$  component of the amide I' band of  $\alpha$ -helical regions and to the amide I' band of unordered regions.<sup>5</sup> A qualitative examination of the spectra suggests that the amount of  $\beta$ -structure decreases in the order: ribonuclease >  $\alpha$ -chymotrypsin > lysozyme > myoglobin, in agreement with the known crystal structure of these proteins.<sup>8-11</sup> The amount of  $\beta$ -structure in ribonuclease is ca. 44%,<sup>9,13</sup> in myoglobin close to zero,<sup>7-13</sup> while the other two proteins fall somewhere in between.

### Proteins of Unknown Conformation

Figure 2 shows the spectra of  $\alpha_s$ -casein,  $\alpha$ -lactalbumin,  $\beta$ -lactoglobulin A, and alkali-denatured  $\beta$ -lactoglobulin A (pD 11.5).  $\alpha_s$ -casein exhibits a single band, centering at  $1646\text{ cm}^{-1}$ , in the amide I' region. This protein is known to assume an apparently random conformation in aqueous solution.<sup>15</sup>  $\alpha$ -lactalbumin, Figure 2b, exhibits an amount of  $\beta$ -structure that appears to be slightly above lysozyme and below  $\alpha$ -chy-

TABLE I  
Fraction of  $\beta$ -Structure in Different Proteins

Protein	X-Ray	CD <sup>a</sup>	ir
Ribonuclease	0.44 <sup>9,17</sup>	0.44 <sup>13,17</sup>	(0.42)
$\alpha$ -chymotrypsin	0.34 <sup>11,17</sup>	0.10 <sup>13,17</sup>	(0.31)
Lysozyme	0.16 <sup>10,17</sup>	0.16 <sup>13,17</sup>	(0.18)
Myoglobin	0.0 <sup>8,17</sup>	0.02 <sup>13,17</sup>	0.0
$\beta$ -lactoglobulin	—	0.45 <sup>19,20</sup>	0.48
$\alpha$ -lactalbumin	—	0.15 <sup>18</sup>	0.18
$\alpha_s$ -casein	—	0.0 <sup>15</sup>	0.0

<sup>a</sup> The value for  $\alpha_s$ -casein is from optical rotatory dispersion measurements.

motrypsin (cf. Figure 1b and c). The amount of  $\beta$ -structure in native  $\beta$ -lactoglobulin appears to be equal to, or higher than in ribonuclease (cf. Figure 1a). Alkali-denatured  $\beta$ -lactoglobulin again shows a single peak in the amide I' region, which centers at 1645  $\text{cm}^{-1}$ , very close to the band of  $\alpha_s$ -casein, and indicates that the  $\beta$ -structure has been completely destroyed.

### Numerical Estimations of $\beta$ -Structure Content

In order to estimate the amount of  $\beta$ -structure in a protein of unknown conformation from the resolved infrared curves, an absorptivity value for the characteristic band at ca. 1630  $\text{cm}^{-1}$  is necessary. (The amount of  $\alpha$ -helix cannot be estimated because the characteristic band almost coincides with the band arising from unordered segments.) The absorptivity for the  $\beta$ -structure band around 1630  $\text{cm}^{-1}$  can be estimated from the resolved spectra of ribonuclease,  $\alpha$ -chymotrypsin, and lysozyme by the relationship:

$$a_\nu = \frac{A_\nu}{c \cdot d \cdot x_\beta} [1 \cdot \text{g}^{-1} \cdot \text{cm}^{-1}]$$

where  $a_\nu$  is absorptivity,  $A_\nu$  absorbance,  $c$  protein concentration in grams per liter,  $d$  cell thickness in centimeters, and  $x_\beta$  the fraction of the protein that is in the  $\beta$ -conformation. Adopting  $x_\beta$  values of 0.44, 0.34, and 0.16 for ribonuclease,  $\alpha$ -chymotrypsin, and lysozyme,<sup>8-11,13</sup>  $a$  values of 4.53, 4.22, and 5.51  $\text{l. g}^{-1} \text{cm}^{-1}$  are obtained. The spread of the values is, unfortunately, rather large. (The use of band areas instead of peak heights leads to no improvement.) The characteristic band concept applies to uniform chains of infinite length only.<sup>16</sup> (The same holds for other spectroscopic methods such as ORD and CD.) The segments of  $\beta$ -structure in the studied proteins of known conformation are relatively short and somewhat distorted.<sup>9-11</sup> The value for lysozyme, the protein with the shortest  $\beta$ -segments, is farthest out of line.

To obtain a semiquantitative idea about the  $\beta$ -structure content of the proteins of unknown conformation, an average absorptivity value of

4.75 l. g<sup>-1</sup> cm<sup>-1</sup> was assumed. Table I compares the obtained values with values calculated from X-ray data<sup>8-11</sup> and circular dichroism data.<sup>13,17-20</sup> (The first three values obtained from infrared data are simply the results of applying the averaged absorptivity to the known structures.) The values for  $\beta$ -lactoglobulin and  $\alpha$ -lactalbumin are reasonably close to estimations from optical rotatory dispersion and circular dichroism data. Taking into account that the secondary structure of globular proteins is not easily described as a mixture of well-defined conformations, the results are probably as good as can be expected. Lysozyme and  $\alpha$ -lactalbumin, which are quite similar in their primary structure, appear to contain equal amounts of  $\beta$ -structure.

### References

1. Miyazawa, T., Shimanouchi, T. & Mizushima, S. (1958) *J. Chem. Phys.* **29**, 611-616.
2. Miyazawa, T. (1962) in *Polyamino Acids, Polypeptides and Proteins*, Stahmann, M. A., Ed., Univ. Wisconsin Press, Madison, pp. 201-217.
3. Bellamy, L. J. (1958) *The Infrared Spectra of Complex Molecules*, 2nd ed., Wiley, New York, pp. 203-232.
4. Krimm, S. (1962) *J. Mol. Biol.* **4**, 528-540.
5. Susi, H., Timasheff, S. N. & Stevens, L. (1967) *J. Biol. Chem.* **242**, 5460-5466.
6. Timasheff, S. N., Susi, H. & Stevens, L. (1967) *J. Biol. Chem.* **242**, 5467-5473.
7. Timasheff, S. N. & Susi, H. (1966) *J. Biol. Chem.* **241**, 249-250.
8. Kendrew, J. C., Watson, H. C., Strandberg, B. E., Dickerson, R. E., Phillips, D. C. & Shore, V. C. (1961) *Nature* **190**, 666-670.
9. Wyckoff, H. W., Tsernoglou, D., Hanson, A. W., Knox, J. R., Lee, B. & Richards, F. M. (1970) *J. Biol. Chem.* **245**, 305-328.
10. Blake, C. C. F., Mair, G. A., North, A. C. T., Phillips, D. C. & Sarma, V. R. (1967) *Proc. Roy. Soc., Ser. B* **167**, 365-377.
11. Blow, D. M. (1969) *Biochem. J.* **112**, 261-268.
12. Hooke, R. & Jeeves, T. A. (1961) *J. Ass. Comput. Mach.* **8**, 212-213.
13. Chou, P. Y. & Fasman, G. D. (1974) *Biochemistry* **13**, 222-245.
14. Susi, H. (1969) in *Structure and Stability of Biological Macromolecules*, Timasheff, S. N. & Fasman, G. D., Eds., Marcel Dekker, New York, pp. 575-663.
15. Herskovits, T. T. & Mescanti, L. (1965) *J. Biol. Chem.* **240**, 639-644.
16. Miyazawa, T. (1960) *J. Chem. Phys.* **32**, 1647-1652.
17. Chen, Y.-H., Yang, J. T. & Martinez, H. M. (1972) *Biochemistry* **11**, 4120-4131.
18. Robbins, F. M. & Holmes, L. G. (1970) *Biochem. Biophys. Acta* **221**, 234-240.
19. Townend, R., Kumosinski, T. F. & Timasheff, S. N. (1967) *J. Biol. Chem.* **242**, 4538-4545.
20. Timasheff, S. N., Townend, R. & Mescanti, L. (1966) *J. Biol. Chem.* **241**, 1863-1870.